Health Informatics

Lecture 1: Introduction

Samantha Kleinberg
samantha.kleinberg@stevens.edu
About me: Samantha Kleinberg

• PhD in CS
• Have worked in biomathematics, bioinformatics/computational biology, biomedical informatics
• Current research
  – Time series data, causal inference
  – Mobile health
  – Decision-making
NICU dataset

• 98 patients with subarachnoid hemorrhage

• Monitoring included
  – Depth and surface EEG
  – Microdialysis
  – Physiologic measurements

(no data on procedures)
Lots of data, but lots of missing data

- Device malfunctions
- Device connected to perform a procedure
And...

• All variables may be missing at once (if measured by single device)
  – Can’t use imputation methods that assume some present values

• Missing values depend on variable + other variables
  – e.g. BG depends on itself, as well as insulin

• Variables are correlated across time
Different data, different physiology

Regulation changes over time

0-96hrs

96-168 hrs
Causes of changes in glucose

Cohort: 17 subjects with T1DM

Sensor data (collected for >72 hours)

- Glucose values
- Insulin dosage
- Activity
- Sleep stage
- Heart rate
- Temperature

With N. Heintzman (UCSD) http://dial.ucsd.edu/what-we-do.php
Results

very vigorous exercise leads to hyperglycemia (fdr < .01) in 5-30 minutes
  – Found using both HR (anaerobic activity zone) and METs

• Not found with strict discretization
• Supported by literature
  (Marliss and Vranic, 2002; Riddell and Perkins, 2006)

Why are you here?
Why health informatics?

**All of Us**<sup>SM</sup> Research Program

**WHAT IS IT?**

**Precision medicine** is a groundbreaking approach to disease prevention and treatment based on people’s individual differences in environment, genes and lifestyle.

The *All of Us* Research Program will lay the foundation for using this approach in **clinical practice**.

https://www.nih.gov/AllofUs-Research-Program
WHAT ARE THE GOALS?

Engage a group of **1 million or more U.S. research participants** who will share biological samples, genetic data and diet/lifestyle information, all linked to their electronic health records. This data will allow researchers to develop more precise treatments for **many diseases and conditions**.

Pioneer a new model of research that emphasizes **engaged research participants, responsible data sharing and privacy protection**.
Take Control of Your Data

You already shed data through time and space all day long, collected by the corporations that track your spending, location, and use of technology. We at The HUMAN Project believe that the people’s data can be better used to serve them, their communities, and society.

What Will We Do?

We will study the lives of 10,000 New York City residents in approximately 4,000 households over the span of decades by collecting measurements across multiple domains and disciplines, and provide this massive, interdisciplinary, and secure dataset to the research community to achieve previously unattainable groundbreaking advancements in medicine, the social sciences, and understanding the true nature of human behavior, while fostering evidenced-based public policies.

Why Now?

Enrollment starting early 2018
And for final projects

If you use MIMIC data or code in your work, please cite the following publication:


https://mimic.physionet.org
Example projects

• Do patients with diabetes with more hospital admissions have more amputations?
• Comparing drug overdose trends to national statistics
• Classification of pneumonia using waveform data
• Difference in ICU stay duration for CHF in T1D vs T2D patients
• Visualization of patient clusters
What could we do with this?

• Who’s the user?
• What problem do they have?
• How can CS help?
But what do we need to know first?

• What’s the data actually? Is it correct?
• Can we really apply ML/DM methods to it? What are the challenges?
• Can we share data freely? Why not?
• How do we preserve privacy

(and much more)
Syllabus

CS 544, Spring 2018

HEALTH INFORMATICS

From medical centers and individual physicians adopting electronic medical records, to patients keeping track of chronic diseases through websites and apps, we live in an era of unprecedented access to health data. These data enable inference of drug side effects, causes of disease, and new treatments, but the new terminologies, policies, and challenges in understanding the data itself can make it difficult for computational researchers to apply their techniques to this new area and for health professionals to begin using informatics to solve practical problems. This course will give both groups the foundation needed to propose, evaluate and develop projects such as secondary analysis of health data and will enable them to begin effective interdisciplinary collaborations. Students will learn how health data is collected (in both hospital and non-hospital settings), how the structure of record systems impacts the research process and interpretation of results, and how to design and evaluate studies involving secondary use of health data (while complying with HIPAA and IRB regulations) in order to gain new medical knowledge and improve healthcare delivery.

Prerequisites None. The course is intended for advanced undergraduate and graduate students from computer science and other disciplines.

Text There is no required textbook. Readings are articles that will be provided.

Evaluation Discussion of the readings is an important part of the course and will count towards the final grade. It is not possible to succeed in this course without participating. There will be a final project (and presentation), as well as one presentation for the journal club assignment and discussions of case studies.

Grades will be: 10% homework, 15% participation, 25% midterm exam, 50% final project.
Syllabus (continued)

CS 544, Spring 2018

Schedule

Below is the weekly schedule of topics, readings for each session, and deadlines. All journal articles will be available electronically through the library or in Canvas. Readings may change, but in that case, an announcement will be made in class.

1/22 Introduction to health informatics. Overview of field, relation to bioinformatics

1/29 Where and why is health data collected? (including meaningful use and PPACA). Including claims data, clinician-generated data, and information exchange.

2/5 What are the data? EMRs, PHRs, and data standards (narrative vs. structured data, ontologies)
  - S. Abhyankar, D. Demner-Fushman, F. M. Callaghan, and C. J. McDonald. Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. *Journal of the American Medical Informatics Association, 1*, 2014 [pdf]

2/12 Ethics in biomedical research. HIPAA, IRBs, and de-identification. We will discuss the rules and regulations governing health research and, from a practical standpoint, how to preserve privacy when conducting research.

2/21 Data reuse and evaluation challenges (what is a diagnosis really?)
  - IRB proposals due

2/26 Research with EMR data, introduction to MIMIC dataset

3/5 Midterm exam
3/19 Methods for evaluation. How can we compare systems and determine if a project is successful?

   • Project proposals due

3/26 Pharmacovigilance and drug discovery

4/2 Journal club discussion of recent papers in health informatics

4/9 Decision Support

4/16 Public health, epidemiology, and mobile health

4/23 Presentations of final projects
   • Written project reports due

4/30 Presentations of final projects
Administrativia

• Course website:  
  http://www.skleinberg.org/teaching/Hi18/

• Prerequisites: none

• Textbook: none, articles

• Everything on syllabus CAN CHANGE (will make announcement in class)

• Workload/grading:
  midterm presentation (25%), final project (50%), participation (15%), homework (10%)
  – Participation includes weekly reading discussions
Key policies

1. No late work. There are few deadlines, but if the deadline’s 11:59pm, work submitted at 12:01am isn’t accepted.
   
   Do not email me saying the time changed as you were submitting. That is the definition of late.
   
   Why? Try submitting an NIH grant or conference paper 2 minutes late!

2. Plagiarism = F
What’s plagiarism?

Presenting someone else’s work as your own

– Copying an entire paper
– Copying parts of other works, without attribution
  • E.g. quotes without citations, images
– Changing a few words, but keeping ideas and structure, without acknowledging source

Easy to avoid! Do your own work and acknowledge all sources. Quotes must be in quotes. Don’t submit a collage.
1. Quotes belong in quotation marks

- **WRONG!**
  - Blah blah. Text from other papers. Blah Blah. [Citation]
  - **Section 1** [Citation]. Text you did not write.
  - **Section 1** [Full text of someone else’s paper] References Else, Someone. “Paper you copied from”

- **Right 😊**
  - My text “Quote from awesome work.” [citation] my text
2. Papers aren’t collages
Seriously, papers are not collages

• “But the other paper said what I wanted to!”
  – Put it in your own words
• “It’s just reference material”
  – Then add a reference to it
• “I copied from your book because you said it so well!”
  – I am not senile and will recognize my own words
3. No copying/close paraphrasing

Don’t Let This Be You!


Plagiarized!

Sloppy Firsts (2001), page 7: “Bridget is my age and lives across the street. For the first twelve years of my life, these qualifications were all I needed in a best friend. But that was before Bridget’s braces came off and her boyfriend Burke got on, before Hope and I met in our seventh grade Honors classes” (McCafferty, quoted in Zhou, “Examples…”).

Opal Mehta (2006), page 14: “Priscilla was my age and lived two blocks away. For the first fifteen years of my life, those were the only qualifications I needed in a best friend. … But that was before freshman year, when Priscilla’s glasses came off, and the first in a long string of boyfriends got on” (Viswanathan, quoted in Zhou, “Examples…”).

“an act of literary identity theft”
(Steve Ross, qtd. in Zhou, “Publisher…”)

http://researchguides.stevens.edu/plagiarism
If you’re having trouble...

• Come to office hours! Monday 2-3 in North 208 and by appointment

• Email me

• Email or meet with CA Ivan Ching (iching@stevens.edu)
What’s health informatics?
Biology

Computer Science:
Theory and practice of computing

Healthcare:
Information science
Storing, processing, etc. information

Biomedical computing

Biomedical computing:
Translational bioinformatics

Bioinformatics

Health informatics
Biomedical informatics
Another definition

- **Definition**: Biomedical informatics (BMI) is the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, driven by efforts to improve human health.

- **Scope and breadth of discipline**: BMI investigates and supports reasoning, modeling, simulation, experimentation, and translation across the spectrum from molecules to individuals and to populations, from biological to social systems, bridging basic and clinical research and practice and the healthcare enterprise.

- **Theory and methodology**: BMI develops, studies, and applies theories, methods, and processes for the generation, storage, retrieval, use, management, and sharing of biomedical data, information, and knowledge.

- **Technological approach**: BMI builds on and contributes to computer, telecommunication, and information sciences and technologies, emphasizing their application in biomedicine.

- **Human and social context**: BMI, recognizing that people are the ultimate users of biomedical information, draws upon the social and behavioral sciences to inform the design and evaluation of technical solutions, policies, and the evolution of economic, ethical, social, educational, and organizational systems.

Main subfields

• (Translational) Bioinformatics
• Public health informatics
• Healthcare/Clinical informatics
• Consumer/mHealth
Example: Translational bioinformatics

Example: Public health Informatics

Fig. 5. Comparison of Twitter self-reports with (a) Venmo payments for alcohol (b) Google Trends for alcohol topics and (c) Google Trends for hookah topics. Dotted lines are at 8am. In (a) the Venmo data is at two-hour resolution for clarity. Venmo data is normalized by all transactions each hour, Google Trends are scaled based on proportion to all searches on all topics, and Twitter posts normalized to all Tweets from the API.

Example: Clinical informatics
Example: Consumer informatics

**TreatYoSelf: Empathy-driven behavioral intervention for marginalized youth living with HIV**

Gabriela Marcu¹, Nadia Dowshen², Shuvaditty Saha¹, Ressa Reneth Sarreal³, Nazanin Andalibi¹

¹College of Computing and Informatics
Drexel University
Philadelphia, PA
{gmarcu, ss3766, naz}@drexel.edu

²Children’s Hospital of Philadelphia
Univ. of Pennsylvania School of Medicine
Philadelphia, PA
dowshenn@email.chop.edu

³University of Maryland,
Baltimore County
Baltimore, MD
saressa1@umbc.edu

**ABSTRACT**

Behavioral intervention technologies are well suited to addressing health behavior such as medication adherence, but only if successfully integrated into a user’s daily life. Little is known about how to design such technologies to be adoptable, adaptable, useful, and feasible in everyday life. We report on the design process for TreatYoSelf, a smartphone application designed to improve medication adherence among youth living with HIV through reminders and positive reinforcement. Using participatory design, our aim was to understand factors related to adoption and acceptance of behavioral intervention technology as part of daily life. Two challenges of living with HIV led to an empathy-driven approach in our design process: (1) HIV is a stigmatized condition, which (2) disproportionately affects the marginalized populations of young African American men who have sex with men and transgender women. We discuss five empathy-driven design strategies: positive and nonjudgmental tone; minimal, avatar-based gamification; motivational and corny messages; nondisclosure through neutral signifiers; and social support through camaraderie. Our approach enabled us to identify and work through factors, often related to stigma and marginalization, which would lead to rejection of TreatYoSelf use in daily life.
AMIA Themes

- **Biomedical Data Visualization**
- **Clinical Informatics**: findings related to the design, development, and implementation of state-of-the-art clinical systems
- **Clinical Research Informatics**: addressing the critical need for effective information management to address the many challenges facing clinical research
- **Clinical Workflow and Human Factors**: human factors aspects of clinical information system implementation and use that revolves around usability, workflow, and patient safety
- **Consumer Informatics and PHRs**: Personal Health Records (PHRs) and the consumer perspective in the use of health information science designed to improve patient engagement, medical outcomes, and the health care decision-making process
- **Mobile Health**: mHealth, Web 2.0, social media, telehealth/telemedicine, and related topics
- **Data Interoperability and Information Exchange**: methods that organizations have undertaken to develop and implement various clinical data integration and exchange activities
- **Data Mining, NLP, Information Extraction Retrieval**: application of data mining, natural language processing, information extraction retrieval to all areas of biomedicine
- **Achieving Meaningful Use**: ways to promote the successful and effective development, implementation, and evaluation of Electronic Health Records as the nation works toward "meaningful use" of these systems
- **Global eHealth**: Global eHealth challenges

- **Informatics Education and Workforce Development**: effort to create a trained HIT workforce
- **Informatics in Health Professional Education**: application of information technology in health professional education
- **Interactive Systems**: human-computer interaction (HCI) research, compelling designs, or innovative interactive technologies
- **Policy and Ethical Issues**: unprecedented national HIT activity and ethical considerations posed as more practitioners and the public interface with these technologies.
- **Public Health Informatics and Biosurveillance**: disease detection, communications, workforce development, standards and interoperability, and best practices to combine the domains of health information science and technology with the practice and science of public health
- **Imaging Informatics**: intersection of imaging science, biomedical engineering and biomedical informatics,
- **Simulation and Modeling**: use of various computer-based simulation and modeling methodologies and tools as they can be applied within the field of biomedical informatics
- **Terminology and Standards Ontologies**: complex issues surrounding standard syntax, semantics, and pragmatics of design, development and use of various application-specific and general purpose clinical terminologies and ontologies
- **Translational Bioinformatics and Biomedicine**: opportunities in biomedical informatics that arise from the storage, retrieval, analysis, and dissemination of molecular and genomic information in a clinical setting context

http://www.amia.org/amia2013/themes
Goals

Conceptual Value/Mission Model for Academic Health Centers

<table>
<thead>
<tr>
<th>Value: Quality</th>
<th>Missions:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient Care</strong></td>
<td><strong>Education</strong></td>
</tr>
<tr>
<td><strong>Right Thing (Vision)</strong></td>
<td>Carriage trade</td>
</tr>
<tr>
<td></td>
<td>Individualization</td>
</tr>
<tr>
<td></td>
<td>Mass customization</td>
</tr>
<tr>
<td></td>
<td>Population-based medicine</td>
</tr>
<tr>
<td></td>
<td>Regional health model</td>
</tr>
<tr>
<td><strong>Right Way (Which, How)</strong></td>
<td>Traditional medicine</td>
</tr>
<tr>
<td></td>
<td>Complementary medicine</td>
</tr>
<tr>
<td></td>
<td>Evidence-based medicine</td>
</tr>
<tr>
<td></td>
<td>Standards of care</td>
</tr>
<tr>
<td></td>
<td>Practice guidelines</td>
</tr>
<tr>
<td></td>
<td>Protocols</td>
</tr>
<tr>
<td></td>
<td>Learning organization</td>
</tr>
<tr>
<td></td>
<td>Person empowerment</td>
</tr>
<tr>
<td></td>
<td>Shared responsibility</td>
</tr>
<tr>
<td><strong>Right Time (When)</strong></td>
<td>Acute care on demand</td>
</tr>
<tr>
<td></td>
<td>Preventive medicine</td>
</tr>
<tr>
<td></td>
<td>Public health education</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Right Person (Who)</strong></td>
<td>Physician</td>
</tr>
<tr>
<td></td>
<td>Integrated staff</td>
</tr>
<tr>
<td></td>
<td>Skill mix</td>
</tr>
<tr>
<td></td>
<td>Patient</td>
</tr>
<tr>
<td></td>
<td>Consumer/public</td>
</tr>
</tbody>
</table>

An Envisioned Cycle That Ties Patient Care with Knowledge Creation and Dissemination

Providers Caring for Patients

Electronic Health Records

Regional and National Public Health and Disease Registries

Biomedical and Clinical Research

Standards for Prevention and Treatment

A “Learning Healthcare System”

Information, Decision-Support, and Order-Entry Systems

Creation of Protocols, Guidelines, and Educational Materials
Personalized medicine

Use individual level characteristics and information to tailor treatment

– Choosing between treatments vs. creating individual treatment
Personalized medicine + informatics

• Using -omics in clinical care
• Decision support
• Find similar patients – treatments, outcomes
Chronic disease

• Diabetes
• Obesity
• Asthma
• Epilepsy

...
Essential data driven feedback loops

Participant self-care
*How is this new medication working for me?*

Clinical care
*How is the patient responding to new care plan?*

Research evidence
*What works best in different contexts?*

http://mobilizelabs.org/why-mhealth
mHealth architecture: Stovepipe versus Open. The narrow waist of the open hourglass will include at least health-specific syntactic and semantic data standards; patient identity standards; core data processing functions such as feature extraction and analytics; and data stores that allow for selective, patient-controlled sharing. Standards should be common with broader health IT standards whenever possible.

Quantified self

January 1
The first entry in my food logbook is
Champagne. Starting the year off right is
easy. I recorded 3,517 food and drink entries
throughout 2012.

December 18
Database is named a semi-finalist in the
NYU Entrepreneurial
Challenge. In April, we were named one of the winners in the
Technology Venture Competition.

December 31
A year of self-tracking is completed.

December 10-17
Finals week and the ITP Winter Show mark the end
of my third term of grad school. My insulin basal
rates are up 60% since the marathon. The work,
stress and lack of exercise have taken their toll.

November 18
I ran the Philadelphia Marathon in
3:47, a new personal record.

November 16
I attended the DiabetesMine Innovation
Summit in San Francisco. A few hours later,
I took a red-eye flight to Philadelphia.

October 28
My worst day of blood sugar control
all year. I stayed at home the entire day,
recovering from yesterday’s run.

October 29
Hurricane Sandy hits New York City.

October 27
My marathon training peaks with
a 20-miles run.

January 9
My friend Peter convinced me
to start using a FitBit, my first
activity tracker. I love the
device, but managed to
break three of them
during the year.

January 13
My second term at NYU’s ITP begins.

January 25 & 28, February 4 & 11
A warm winter meant I was able to do 10-14-
mile long weekend runs with my friend Ryan.
My basal insulin rates dropped 20% as a result.

February 13
My best day of the year for
diabetes control, with blood sugars in range 99% of the day.

March 12
My first A1c blood test of
2012 is 5.6, my best result ever.

http://databetic.com
Self-tracking helped me achieve the best diabetes control of my life, as evident in my A1c results. This blood test is the standard metric for measuring diabetes control. The lower the reading the better. For context, I’ve included 10 years worth of my A1c readings.
Exercise

Training for a Marathon

Most exercise lowers a patient’s blood sugar right away. Frequent exercise can also lower overall insulin requirements.

In 2012, I trained for and ran the Philadelphia Marathon on November 18. During the year, I also had a few periods where I ran very little. These fluctuations translated into significant changes in my insulin levels throughout the year.

1. The winter was unusually warm. I began doing long weekend runs of 10-14 miles with my friend Ryan in late January. My insulin rates dropped about 20% as a result. For the rest of the spring, my insulin stayed steady.

2. At the end of my grad school term, my mileage dropped considerably. Insulin rates went up about 30% as a result.

3. Marathon training peaked in late at some of the lowest levels of the summer, I slowly got it. My insulin rates had dropped at signed up for the Philadelphia & July.

4. By the time of the marathon, my insulin rates rose a lot. I barely ran at all. It was also a t end of my term at grad school, requirements. Not until the holiday allowing my insulin rates to retu
Challenges/open problems

• Data
• Replication/validation
• Privacy/sharing
• Translation
Challenge: replication

Table 1: Reproducibility of research findings
Preclinical research generates many secondary publications, even when results cannot be reproduced.

<table>
<thead>
<tr>
<th>Journal impact factor</th>
<th>Number of articles</th>
<th>Mean number of citations of non-reproduced articles*</th>
<th>Mean number of citations of reproduced articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;20</td>
<td>21</td>
<td>248 (range 3–800)</td>
<td>231 (range 82–519)</td>
</tr>
<tr>
<td>5–19</td>
<td>32</td>
<td>169 (range 6–1,909)</td>
<td>13 (range 3–24)</td>
</tr>
</tbody>
</table>

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.


Drug development: Raise standards for preclinical cancer research
C. Glenn Begley & Lee M. Ellis
Nature 483, 531–533 (29 March 2012) doi: 10.1038/483531a
Challenge: Privacy

"87% of the U.S. Population are uniquely identified by {date of birth, gender, ZIP}.

http://latanyasweeney.org/work/identifiability.html
Challenge: Alert fatigue

![Graph showing referral rate over weeks]

Embi P J, and Leonard A C J Am Med Inform Assoc doi: 10.1136/amiajnl-2011-000743
For next week

• Be sure to read Wicks et al. paper in syllabus!
• Response due Friday by noon!